Hausman Tests for Inefficient Estimators: Application to Demand for Health Care Services

Michael Creel*

February, 2003

Abstract

The Hausman (1978) test is based on the vector of differences of two estimators. It is usually assumed that one of the estimators is fully efficient, since this simplifies calculation of the test statistic. However, this assumption limits the applicability of the test, since widely used estimators such as the generalized method of moments (GMM) or quasi maximum likelihood (QML) are often not fully efficient. This paper shows that the test may easily be implemented, using well-known methods, when neither estimator is efficient. To illustrate, we present both simulation results as well as empirical results for utilization of health care services.

Keywords: Hausman test, specification testing, health care utilization

JEL codes: C12, C52, I10

*Department of Economics and Economic History, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona) Spain. Comments are welcome. Email: michael.creel@uab.es. Tel. 93-581-1696.
This research was supported by grants SEC 2000-0390 and SGR2001-0164.

1 Introduction

The Hausman test (Hausman, 1978) is based on the idea that the difference between two consistent estimators tends to zero. One of the estimators, say $\hat{\theta}_1$, is consistent under the null of correct specification, but inconsistent under the alternative. The other estimator, say $\hat{\theta}_2$, is consistent under both the null and the alternative hypotheses. Under the alternative hypothesis of misspecification, $\hat{\theta}_1$ will no longer be consistent, but $\hat{\theta}_2$ will remain so. In this case the difference vector $\Delta = \hat{\theta}_2 - \hat{\theta}_1$ will have a nonzero probability limit, which will cause the test statistic to ultimately reject the null of correct specification.

Hausman's paper, and most subsequent research, has concentrated on the case that $\hat{\theta}_1$ is fully efficient when the associated model is correctly specified. In this case, the asymptotic variance of Δ, V_Δ, is simply the asymptotic variance of the inefficient estimator minus that of the efficient estimator. With this simplification, the test is easily implemented using consistent estimators of the two asymptotic variances.

This paper addresses the case where neither of the two estimators is fully efficient. For the purposes of this paper, it bears emphasizing that widely used estimators such as quasi maximum likelihood (QML) and the generalized method of moments (GMM) are not fully efficient, in general. When performing a Hausman test comparing two inefficient estimators, V_Δ will involve the asymptotic covariance of the estimators. It will not cancel out as in the case analyzed by Hausman.

Some previous research has examined special cases of pairs of inefficient estimators. Ruud (1984) considers the case where a likelihood function may be factored as the sum of two likelihoods. The obvious example is a split sample. When observations are independent, the sample may be split into two equal parts, and each part may be used separately to calculate the two estimators. Sample splitting leads to an important loss of power, since the asymptotic variances of the estimators are twice as large as when the full sample is used. Nor is it a general solution, since it will not solve the problem of correlation between the estimators when the observations are not independent of one another. The idea of sample splitting has been further pursued in the context of QML estimation by de Luna and Johansson (2001). Newey (1985) also presents Hausman tests based on inefficient estimators. He considers GMM estimators defined by different linear combinations of a *given* vector of moment conditions. The major limitation of these authors' results is that their frameworks are very specific and are highly restrictive in terms of the pairs of estimators that may be considered.

In the empirical literature, Windmeijer and Santos Silva (1997) perform a Hausman test based on Poisson QML and GMM estimators. Since neither the QML nor the GMM estimator is efficient, the authors use a split sample Hausman test, following Ruud (1984) and Browning and Meghir (1991). In their application, Windmeijer and Santos Silva find that this version of the test does not reject exogeneity. One wonders whether the reason for non-rejection might be the loss of power entailed by splitting the sample.

This paper re-examines the Hausman test when neither estimator is efficient. We find that the standard full sample Hausman test is simple to apply in the general case, since accounting for the covariance of the estimators is not at all difficult, and may be done using standard methods. The result is simple, but it appears to have been overlooked, and it does have practical importance. The result also immediately suggests some closely related tests. Simulation results are provided that show that the standard Hausman test may be misleading when it is calculated using a pair of inefficient estimators, and that the modified test performs properly under the same circumstances. We apply the test to data on demand for health care visits and private health care insurance. We find strong evidence that health care insurance is an endogenous variable in explaining demand for certain forms of health care.

In the next section we present the main results of the paper. The third section presents the simulation results and the fourth section contains the empirical results.

2 Theory

Hausman (1978) recognized from the beginning that the essential problem in basing a test on estimators that may not be efficient is that the covariance between the estimators must be estimated. In this section we show that methods that have since become standard allow this covariance to be estimated without difficulty. The argument is a straightforward application of existing results on extremum estimators, so results are stated without proof or explicit regularity assumptions. Gallant (1987), Chapter 3, for example, gives regularity conditions and formal proofs of the results that are used here. Newey and West (1987a) and Davidson and MacKinnon (1993, pp. 616-619) present results that are suggestive of those presented here, but which do not explicitly discuss the Hausman test.

An extremum estimator $\hat{\psi}$ may be defined as

$$\hat{\psi} = \arg\max_{\psi \in \Psi} s_n(\psi, Z_n),$$

where Z_n is the data. Extremum estimators encompass minimum distance estimators, defined in terms of

$$s_n(\psi, Z_n) = \frac{1}{n} \sum_{t=1}^{n} s_t(\psi, z_t),$$

where z_t is the data for one observation, and method of moments estimators, defined in terms of

$$s_n(\psi, Z_n) = m_n(\psi, Z_n)' W_n m_n(\psi, Z_n),$$

where $E_\psi m_n(\psi, Z_n) = 0$ and W_n is a matrix with a finite, positive definite almost sure limit. An extremum estimator will, given some regularity conditions, have an almost sure limit, say ψ_0, and will be asymptotically normally distributed:

$$\sqrt{n}\,(\hat{\psi} - \psi_0) \xrightarrow{d} N\,[0, \mathcal{V}_\infty]$$

where $\mathcal{V}_\infty = \mathcal{J}_\infty(\psi_0)^{-1} \mathcal{I}_\infty(\psi_0) \mathcal{J}_\infty(\psi_0)^{-1}$. Here, $\mathcal{J}_\infty(\psi_0)$ is the almost sure limit of

$\mathcal{J}_n(\psi_0) = \frac{\partial^2}{\partial\psi\partial\psi'} s_n(\psi)|_{\psi_0}$, and $I_\infty(\psi_0) = \lim_{n\to\infty} Var\sqrt{n} g_n(\psi_0)$, where we use the notation $g_n(\psi_0) = \frac{\partial}{\partial\psi} s_n(\psi)|_{\psi_0}$.

To test a hypothesis of the form $H_0 : R\psi_0 = 0_{(q\times 1)}$, the Wald statistic

$$W = n\hat{\psi}' R' \left(R\widehat{V}R'\right)^{-} R\hat{\psi}, \tag{1}$$

where $\widehat{V}$ is a consistent estimator of V_∞, is asymptotically distributed as a $\chi^2(r)$ random variable, where r is the rank of $RV_\infty R'$.

Now, turning to the Hausman test, we assume that both estimators are extremum estimators, defined by

$$\hat{\theta}_1 \equiv \arg\max_{\theta\in\Theta} s_n^1(\theta, Z_n) \tag{2}$$

$$\hat{\theta}_2 \equiv \arg\max_{\theta\in\Theta} s_n^2(\theta, Z_n), \tag{3}$$

where $\Theta \subset \Re^k$. If we define $\psi = \begin{pmatrix} \theta_1' & \theta_2' \end{pmatrix}' \in \Theta \times \Theta$, it is clear that the omnibus estimator

$$\hat{\psi} \equiv \arg\max_{\psi\in\Theta\times\Theta} s_n(\psi) = s_n^1(\theta_1, Z_n) + s_n^2(\theta_2, Z_n) \tag{4}$$

will lead to the same values for the estimators as in equations 2 and 3. That is, $\hat{\psi} \equiv \left(\hat{\theta}_1', \hat{\theta}_2'\right)'$. It bears noting that the omnibus estimator is an extremum estimator, and the theory reviewed above will apply, given regularity conditions. Define $\psi_0 = (\theta_A', \theta_0')'$ as the almost sure limit of $\hat{\psi}$. Under the null hypothesis of correct specification, so that both estimators are consistent, $\| \theta_A - \theta_0 \| = 0$, so $R\psi_0 = 0_{(k\times 1)}$. When the first estimator is inconsistent due to misspecification, $\| \theta_A - \theta_0 \| \neq 0$.

Now, if the dimension of θ is k, say, then we can define the matrix

$$R = \begin{bmatrix} I_k & -I_k \end{bmatrix}. \tag{5}$$

When $\hat{\theta}_1$ is asymptotically efficient, the Wald test, in equation 1, of the restriction $R\psi = 0$ is asymptotically equal to the standard Hausman test, and it *is* the Hausman test if the variance estimator that is used, $\hat{V}$, incorporates the information that $\hat{\theta}_1$ is efficient.[1] The advantage of putting the problem into the framework of extremum estimation is that it is immediately clear how a Hausman-type test may be performed when neither of the estimators is fully efficient. Since the two sub-objective functions that define the omnibus estimator share no parameters, the Hessian matrix $J_\infty(\psi_0)$ is simply the block-diagonal matrix formed by the limiting Hessians of the separate estimators defined by equations 2 and 3:

$$J_\infty(\psi_0) = \begin{bmatrix} J^1_\infty(\theta_A) & 0_{k\times k} \\ 0_{k\times k} & J^2_\infty(\theta_0) \end{bmatrix}.$$

These components may be estimated as usual. The matrix $I_\infty(\psi_0)$ may be written as

$$I_\infty(\psi_0) = \begin{bmatrix} I^1_\infty(\theta_A) & I^{12}_\infty(\psi_0) \\ \cdot & I^2_\infty(\theta_0) \end{bmatrix}.$$

Without full efficiency of the first estimator, the off-diagonal covariance term will not cancel out of the test statistics as it does in the standard case, and it will be necessary to define a consistent estimator. While the on-diagonal blocks may be estimated by whatever means are appropriate given the way the estimators are defined, we will discuss means of estimating the entire matrix.

Recall that

$$\begin{aligned} I_\infty(\psi_0) &= \lim_{n\to\infty} Var\sqrt{n}g_n(\psi_0) \\ &= \lim_{n\to\infty} nE\left\{g_n(\psi_0)g_n(\psi_0)'\right\} - \lim_{n\to\infty} nE\left\{g_n(\psi_0)\right\}E\left\{g_n(\psi_0)'\right\} \\ &\equiv A_\infty(\psi_0) - B_\infty(\psi_0). \end{aligned}$$

In general, it is not possible to estimate $B_\infty(\psi_0)$ consistently. However, it is commonly the case that $B_\infty(\psi_0) = 0$. For example, this will hold for a minimum distance estimator

[1] The asymptotic efficiency of $\hat{\theta}_1$ implies that the asymptotic covariance between $\hat{\theta}_1$ and $\hat{\theta}_2$ is equal to the asymptotic variance of $\hat{\theta}_1$ (see Hausman, 1978, Lemma 2.1) . This is what causes the asymptotic covariance to cancel out of the formula for the standard Hausman test.

if $E_\psi g_t(\psi) = E(D_\psi s_t(\psi) = 0, \forall$t. This will be the case for QML estimators when the data are independently and identically distributed, for example. It also holds for a method of moments estimator if the moment conditions are of the form $m_n(\psi, Z_n) = \frac{1}{n}\sum_{t=1}^{n} m_t(\psi, z_t)$ and $E_\psi m_t(\psi, z_t) = 0, \forall t$. We will assume henceforth that we are in a situation such that $B_\infty(\psi_0) = 0$.

With this, for minimum distance estimators,

$$I_\infty(\psi_0) = \lim_{n\to\infty} Var \frac{1}{\sqrt{n}} \sum_t g_t(\psi_0) \equiv \Omega_\infty(\psi_0),$$

and for method of moments estimations,

$$\begin{aligned} I_\infty(\psi_0) &= \lim_{n\to\infty} Var \left\{ 2M_n(\psi_0) W_n \frac{1}{\sqrt{n}} \sum m_t(\psi_0) \right\} \\ &\equiv 4M_\infty(\psi_0) W_\infty \Xi_\infty(\psi_0) W_\infty M_\infty(\psi_0)', \end{aligned}$$

where $M_n(\psi) = D_\psi m_n(\psi)'$, $M_\infty(\psi_0)$ is its almost sure limit, evaluated at ψ_0, and

$$\Xi_\infty(\psi_0) = \lim_{n\to\infty} Var \frac{1}{\sqrt{n}} \sum_t m_t(\psi_0). \tag{6}$$

The remaining problem, as the case may be, is the estimation of $\Omega_\infty(\psi_0)$ or $\Xi_\infty(\psi_0)$. These matrices are the asymptotic covariances of vector valued processes. A number of estimators are available. With dependent observations, the estimators of Newey and West (1987b), Gallant (1987, pg. 533), and Andrews and Monahan (1992) are possibilities. With independent observations

$$\widehat{\Omega_\infty(\psi_0)} = \frac{1}{n}\sum_{t=1}^{n} g_t(\hat{\psi}) g_t(\hat{\psi})' \tag{7}$$

$$\widehat{\Xi_\infty(\psi_0)} = \frac{1}{n}\sum_{t=1}^{n} m_t(\hat{\psi}) m_t(\hat{\psi})' \tag{8}$$

will provide consistent estimators, in many cases.

We have now seen how to implement the standard Hausman test when neither estimator is efficient. But the above suggests two additional ways to test correct specification, one which is well-known, and the other which is new. For clarity, the presentation will be done in terms of GMM estimators. Consider the omnibus estimator when both estimators can be put in the GMM form. In this case we can define the two GMM estimators as

$$\hat{\theta}_1 \equiv \arg\max_{\theta\in\Theta} s_n^1(\theta, Z_n) = m_n^1(\theta_1, Z_n)' W_n^1 m_n^1(\theta_1, Z_n), \tag{9}$$

$$\hat{\theta}_2 \equiv \arg\max_{\theta\in\Theta} s_n^2(\theta, Z_n) = m_n^2(\theta_2, Z_n)' W_n^2 m_n^2(\theta_2, Z_n). \tag{10}$$

We can define the omnibus moment condition $m_n(\psi) = \begin{bmatrix} m_n^1(\theta_1, Z_n)' & m_n^2(\theta_2, Z_n)' \end{bmatrix}'$, and the omnibus GMM estimator

$$\hat{\psi} \equiv \arg\max_{\psi\in\Theta\times\Theta} s_n(\psi) = m_n(\psi)' \begin{bmatrix} W_n^1 & 0_{pq} \\ 0_{qp} & W_n^2 \end{bmatrix} m_n(\psi), \tag{11}$$

where p is the number of moment conditions that defines $\hat{\theta}_1$ and q is the number of moment conditions that defines $\hat{\theta}_2$. That is, the omnibus estimator also has a GMM representation. Now, since the estimators will be correlated when neither is efficient, it is clear that the weighting matrix that defines this (equation 11) GMM estimator is not the efficient weight matrix, even if W_n^1 and W_n^2 are the efficient weight matrices for the two separate estimators. A different Hausman test may be based on the omnibus GMM estimator that uses the overall efficient weight matrix (the inverse of a consistent estimator of the overall covariance of the moment conditions, in equation 6). One may use the Wald test for the hypothesis $\theta_1 = \theta_2$. This test should be more powerful than the usual Hausman test, since it is based upon a more efficient estimator.

Finally, the previous test is based upon the unrestricted estimator, where the two parameter vectors are not restricted in estimation. If we take the previous moment condition but impose the restriction that the two estimators be equal, then we have $\psi = \theta$, and the moment condition becomes $m_n(\psi) = \begin{bmatrix} m_n^1(\psi, Z_n)' & m_n^2(\psi, Z_n)' \end{bmatrix}'$. This leads to an new overidentified GMM estimator, based upon pooling the moments that define the separate estimators to define a single estimator. Now we may apply GMM using the optimal weighting matrix, estimated by whatever means are appropriate. We have overidentification, and standard results tell us that the GMM criterion test statistic $ns_n(\hat{\psi})$ is asymptotically central chi-square with $p+q-k$ degrees of freedom when the moment conditions are correctly specified. The GMM criterion test statistic is essentially a score test applied to the omnibus model, while the Hausman test is a Wald test.

In this section we have defined three test statistics. The first is the standard Hausman test, but with the covariance between the estimators taken into account at the point of testing, but not when estimating. This can be thought of as a Wald test applied to an inefficient GMM estimator. We will refer to this as the H1 test. The second is the same test, but using the covariance between the estimators to improve the efficiency of estimation. This will be referred to as the H2 test. The final GMM criterion test with pooled moments and the restriction imposed will be referred to as the CRIT test. The original, uncorrected Hausman test, which is not asymptotically valid in the general case, will be referred to as the H0 test.

The H1 and H2 tests are Wald tests applied to unrestricted GMM estimators. Burnside and Eichenbaum (1996) find that the true size of such Wald tests often exceeds the nominal size, especially when a number of restrictions are tested jointly. They find that using a covariance estimator calculated using the formulae for the unrestricted estimator, but evaluated at the restricted estimator (that used for the CRIT test) can improve the small sample performance of Wald tests. In what follows we will report results based upon covariance estimators calculated using both the unrestricted and restricted estimators.

Finally, a Hausman test may be based upon the entire vector of differences, or on a sub-vector, where some of the rows of the matrix R in equation 5 are dropped. In what follows we will present results both for the full version where equality of all parameters is tested, and a version of the test where equality of a single parameter is tested. To explain the notation that is used to describe the tests, an "f" means that the full set of restrictions that all parameters are equal is tested, while an "s" means that a single restriction is tested. The notation "be" means that the Burnside-Eichenbaum suggestion for estimating the covariance is used, otherwise the standard estimator for a Wald test is used. For example, "H2(sbe)" means the H2 test of a single restriction, with the Burnside-Eichenbaum suggestion. H0(f) is the standard Hausman test, using all restrictions.

3 Simulations

In this section we examine two simple situations that illustrate the problems one may encounter when using the standard Hausman test with inefficient estimators, and that examine the performances of the alternative tests.

3.1 A linear model

Consider a linear model with heteroscedastic errors and a potentially endogenous regressor, generated by random sampling from the following model:

$$
\begin{aligned}
y &= 0 + z + \eta\sigma \\
\begin{bmatrix} z \\ \eta \\ w \end{bmatrix} &\sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \cdot & 1 & 0 \\ \cdot & \cdot & 1 \end{bmatrix} \right) \\
\sigma &= (1 + z)
\end{aligned}
$$

Thus, z will be an endogenous regressor when $\rho_1 \neq 0$. We may use w as an instrument for z. Since the errors are heteroscedastic, the OLS estimator will not be efficient (it is a QML estimator), and the standard Hausman will not be valid. It can be shown that the test statistics are invariant to a scalar multiple of σ.

To investigate size, we set $n = 100, \rho_1 = 0, \rho_2 = 0.5$, and we perform 10,000 replications.[2] Since $\rho_1 = 0$, z is exogenous. We use the OLS and IV estimators of the model

[2]All results in this paper were obtained using GNU Octave (www.octave.org). All data and estimation programs needed to replicate the results in this paper are available upon request from the author.

$y = \beta_1 + \beta_2 z + \varepsilon$ to calculate the test statistics. The instruments used to calculate the IV estimator are $\{1, w, w^2\}$. The full ("f") tests are based upon equality of the OLS and IV estimators of both $\beta_1$and β_2, while the single ("s") tests check equality of β_2, the coefficient of the potentially endogenous regressor. Table 1 presents the results for the various tests. We see that the standard Hausman test (H0) has serious size distortions in both the full and single restriction versions of the test. The CRIT test performs quite well. The H1 and H2 tests perform quite well, except for the H1(f) and H2(f) tests, which under-reject. The "be" versions of the H1 and H2 tests all have true size close to nominal size.

To examine power, we repeat the simulation, with everything as above but setting $\rho = 0.3$. The results are reported in Table 2. Of the tests that were found to have proper size, the H2(sbe) test is the most powerful, though the H1(sbe) test is a close second.

3.2 A count data model with a latent variable

Consider a count data model with a normally distributed latent variable that is potentially correlated with an observed regressor, generated by random sampling from the following model:

$$
\begin{aligned}
\begin{bmatrix} z \\ \eta \\ w \end{bmatrix} &\sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \cdot & 1 & 0 \\ \cdot & \cdot & 1 \end{bmatrix} \right) \\
y &\sim \text{Poisson}(\lambda) \\
\lambda &= e^{z+\eta-1/2}
\end{aligned}
$$

The two estimators used to perform the Hausman test are the Poisson QML estimator that is defined assuming $y \sim \text{Poisson}(\lambda)$ where $\lambda = \exp(\beta_1 + \beta_2 z)$ (*i.e.*, the latent variable is ignored) and the nonlinear instrumental variables (NLIV) estimator suggested by Mullahy (1997), which uses the residual function $\exp(-\beta_1 - \beta_2 z)y - 1$ and the instruments $\{1, w, w^2\}$. Again, z will be an endogenous regressor when $\rho_1 \neq 0$, so the Poisson QML estimator will not be consistent in this case. The IV estimator is consistent. Neither estimator is efficient in either case.

First we set $n = 500, \rho_1 = 0, \rho_2 = 0.5$. In this case the latent variable is uncorrelated with the regressor, so both the NLS and the NLIV estimators are consistent. We use 10,000 replications. Table 3 presents rejection frequencies for the 10%, 5% and 1% significance levels, respectively. We see that the standard (H0) test presents very serious size distortions in both the full and single restriction version. The CRIT test also seriously over-rejects. The "f" versions of the H1 and H2 tests also seriously over-reject, and the "fbe" versions over-reject, too, though not as seriously. The H1(s), H2(s) and H2(sbe) tests all have true size almost equal to nominal size. The "be" version slightly under-rejects, while the other two over-reject slightly.

Next, to check power, we repeat the above scenario, but setting $\rho_1 = 0.3$, so that z is endogenous. Table 4 reports the results. The H1(s) and H2(s) are more powerful than the H2(sbe) test. This is perhaps expected, since the results on size indicate that the H2(sbe) test is the most conservative of the three. We do not comment on the power of the other tests, since they were found to have serious size distortions.

In summary, these simulation results show that the standard Hausman test can suffer from serious size problems when neither of the estimators it is based upon is efficient. The CRIT test for the overidentifying restrictions of the pooled GMM estimator also can be seriously distorted. Of the tests proposed in this paper, the single restriction tests appear to be quite reliable, especially when the Burnside-Eichenbaum covariance estimator is used.

4 Demand for health care and insurance coverage

Much research effort has investigated the determinants of health care usage, with a small sample of papers being Pohlmeier and Ulrich (1995); Gurmu (1997) and Deb and Holmes (2000). Variables such as private insurance coverage or self-reported health status may be jointly determined with variables related to usage of health care services

(Cameron *et. al.*, 1988; Windmeijer and Santos Silva, 1997; Vera-Hernández, 1999). For example, if both usage and the decision to purchase private insurance are in part determined by an unobservable personal characteristic such as health status, then there will exist a problem of endogeneity in the estimation of the usage model, if usage depends upon insurance status.

Currently, the most common means of estimating a model for count data while taking into account endogeneity is apply a GMM estimator (Windmeijer and Santos Silva, 1997; Mullahy, 1997; Terza, 1998; Vera-Hernández, 1999). Another possibility is to estimate a bivariate model for both endogenous variables by maximum likelihood (ML), using a sufficiently flexible bivariate density (Terza, 1998; Van Ophem, 2000, Romeu and Vera-Hernández, 2001). This idea is incompletely developed at present, due to the difficulties involved in finding a computationally tractable bivariate density that is sufficiently flexible to warrant the assumption of correct specification. Under the more traditional approach, when facing the choice between using an estimator that ignores endogeneity and a GMM estimator that accounts for it, a Hausman test of the type presented in this paper will be a useful tool. The modified version of the Hausman test will be needed when the estimator that is based upon the assumption of exogeneity cannot be assumed to be a ML estimator, perhaps because of unmodeled latent variables or other reasons.

4.1 Data

We use the 1996 Medical Expenditure Panel Survey (MEPS) data, which contains six different measures of annual health care usage[3]. These are office-based doctor visits (OBDV), outpatient visits (OPV), emergency room visits (ERV), inpatient visits (IPV), dental visits (DV), and number of prescription drugs taken (PRESCR). In order to obtain a simple model that can pass specification tests, we limit the sample to people of age between 40 and 65 years, inclusive, and we estimate separate models for men and women. The explanatory variables are months of private insurance coverage during

[3]The raw data (file HC-012) is available at www.meps.ahrq.gov, and the programs used to prepare the data, the prepared data, and the estimation routines are available from the author.

the year (PRIV), months of public insurance coverage during the year (PUB), age (AGE), years of schooling (EDUC), and family income (INC). All the variables with the exception of INC are directly available. INC was constructed by summing the incomes of all members of the family. Observations for which any family member's income was "hot decked" were dropped.[4] There are 984 observations (men) and 1117 observations (women) for which all needed variables are available.

We do not condition on any measure of health status. The health status measures available in the MEPS data are either perceived health status, sometimes self-reported, sometimes reported by other family members, and objective measures that are quite specific and that may not be good indicators of overall health. As such, health status is treated as a purely latent variable. The fact that both health care usage and private insurance status are likely to depend upon health status is the reason that one suspects endogeneity of private insurance status in a model of health care usage. We assume that public insurance status is exogenous in a model of health care usage. While there may be some grounds for questioning this assumption, it appears to be reasonable for this data set in light of the specification test results reported below.

Let η be a latent index of health status that has expectation equal to unity.[5] We suspect that η and *PRIV* may be correlated, but we assume that η is uncorrelated with the other regressors. For each of the health care usage measures, represented as y in the following equation[6], we assume that

$$\begin{aligned} & E(y|PUB, PRIV, AGE, EDUC, INC, \eta) \\ &= \exp(\beta_1 + \beta_2 PUB + \beta_3 PRIV + \beta_4 AGE + \beta_5 EDUC + \beta_6 INC)\eta. \end{aligned}$$

[4]"Hot decking" is a term used in the MEPS documentation to describe a method of replacing missing data with conditional or unconditional means of the variable. See the documentation for the HC-012 file, available at www.meps.ahrq.gov, for more details.

[5]A restriction of this sort is necessary for identification.

[6]The regression coefficients are assumed to vary according to the usage measure, but this is suppressed in the notation for readability.

We use the Poisson QML estimator of the model

$$\begin{aligned} y &\sim \text{Poisson}(\lambda) \\ \lambda &= \exp(\beta_1 + \beta_2 PUB + \beta_3 PRIV + \beta_4 AGE + \beta_5 EDUC + \beta_6 INC). \end{aligned} \tag{12}$$

Since much previous evidence indicates that health care services usage is overdispersed[7], this is almost certainly not an ML estimator, and thus is not efficient. However, when η and $PRIV$ are uncorrelated, this estimator is consistent for the β_i parameters, since the conditional mean is correctly specified in that case.

When η and $PRIV$ are correlated, Mullahy's (1997) NLIV estimator that uses the residual function

$$\varepsilon = \frac{y}{\lambda} - 1,$$

where λ is defined in equation 12, with appropriate instruments, is consistent. As instruments we use all the exogenous regressors, as well as the cross products of PUB with the variables in $Z = \{AGE, EDUC, INC\}$. That is, the full set of instruments is

$$W = \{1 \quad PUB \quad Z \quad PUB \times Z \}.$$

Since PUB is rather strongly negatively correlated with $PRIV$ ($\rho = -0.485$), and since the coefficient of determination when $PRIV$ is regressed by ordinary least squares on the instruments, W, is $R^2 = 0.40$, we conclude that the instruments are reasonably strong. There are 8 instruments and 6 parameters to estimate.

Neither the the QML nor the NLIV estimators are efficient, which suggests that the standard Hausman test may give misleading results. In order for the results of any Hausman-type test to be convincing, we should have evidence that the NLIV estimator is in fact consistent. To check the correctness of the specification of the conditional mean and the validity of the instruments, we put the NLIV estimator in the GMM form, and report the omnibus specification test $ns_n(\hat{\theta})$ where $s_n(\hat{\theta})$ is the GMM cri-

[7] Overdispersion exists when the conditional variance is greater than the conditional mean. If this is the case, the Poisson specification is not correct.

terion function using the optimal weighting matrix. Table 5 presents the marginal significance level at which the null hypothesis of correct specification may be rejected, for each of the use measures, and for men and women. In only one case do we reject at a level below 10%. Given that this sort of test often over-rejects in finite samples (Hansen and Heaton, 1996), we conclude that the fairly simple model seems to be an adequate specification of the conditional mean.

When testing exogeneity of private insurance status, we hypothesize that endogeneity is most likely to be present when the level of use of a type of care is to a large degree under the control of the patient. Office-based visits (OBDV) and dental visits (DV) seem to be the two clearest cases. While the practitioner certainly can influence the number of visits of these types, the patient also has a good degree of control, since the patient initiates the visits. On the other hand, outpatient visits (OPV) and inpatient visits (IPV) require a physician's intervention for a usage event to occur. Emergency room visits are due to accidents or unexpected illnesses that are (at least usually) severe enough that immediate care is necessary, and are thus unlikely to be strongly influenced by private insurance status. The number of prescription drugs taken (PRESCR) is an unclear case, since a physician must prescribe the drugs, but the patient can initiate visits with a number of physicians. In sum, we expect that endogeneity may be a problem when analyzing the OBDV and DV measures of use. Private insurance status seems unlikely to be endogenous in models for OPV, IPV and ERV. We have no strong prior opinion in the case of PRESCR.

Tables 6 and 7 present the marginal significance levels (p-values) at which exogeneity of *PRIV* may be rejected, for all of the tests and for each of the usage variables, for men and women, respectively. The simulation results in Section 3.2, which are for a model similar to that estimated here, suggest that the most reliable test is the H2(sbe) tests, so those are the results we focus on for the purpose of testing exogeneity. All of the single restriction tests are calculated using the difference of the estimated coefficients of PRIV.

In Tables 6 and 7, the H2(sbe) test gives strong evidence that private insurance status is endogenous in the cases of the OBDV and PRESCR usage measures, for both men and women. The case of DV, for men, is the only other measure for which the H2(sbe) p-value approaches conventional levels for rejection. The other tests that were found to work reasonably well in the simulations (the H1(s), H1(sbe), and H2(s) tests) all give similar results, except that exogeneity is questionable in the case of IPV for men. It is worth noting that the conventional Hausman test gives a p-value of one in a number of cases. This is because the test statistic took on a negative value in those cases. This occurs since the difference of the estimated covariance matrices of the GMM and GML estimators is not necessarily positive semidefinite. None of the different versions of the H1 and H2 tests are affected by this problem, since taking the covariance between the estimators into account causes the overall estimated covariance matrix to be positive semidefinite.

In summary, private insurance status appears to be endogenous for the OBDV and PRESCR usage measures, for both men and women. This is not unexpected in the case of OBDV, but the case of PRESCR is perhaps surprising. The unobserved factors that lead to higher than average consumption of prescription drugs appear to be correlated with those that lead to seeking private insurance coverage. The effect of unobserved health status could very plausibly explain this phenomenon. The factors that lead physicians to prescribe drugs could also be of some importance.

5 Conclusions

This paper has presented several modified versions of the Hausman test that may be used when neither of the estimators is efficient. The standard Hausman test is not valid in this case, as has been illustrated using simulation. The simulation results illustrate the fact, already known in the literature (*e.g.* Hansen and Heaton, 1996), that the CRIT test, while asymptotically valid, can suffer from serious size distortions. The simulations, within their limited scope, show that the single restriction versions of the H1 and H2 tests performs quite well. We do not attempt a broader Monte Carlo study simple since the number possibilities to investigate is overwhelming. When we turn to an empirical investigation, we find that the new tests gives results that are plausible, given prior beliefs regarding the determinants of health care visits.

An empirical result that should be highlighted is that, for health care usage, exogeneity seems to be a fairly innocuous assumption in certain situations. When this is the case, univariate maximum likelihood methods, which have been extensively developed in recent research, may be safely used, with the resulting efficiency. On the other hand, the results strongly indicate that endogeneity is a problem for certain usage measures. In these cases GMM estimation seems to be the best alternative, at least until the multivariate ML approach is better developed. In this study, public insurance coverage is available, and provides a strong instrument for private insurance coverage. In other studies, for example, that of Vera-Hernández (1999), where all individuals have public coverage, such an instrument is not available, and the GMM estimates suffer from imprecision. This is the sort of situation that might motivate additional work on flexible multivariate densities, in order to deal with endogeneity while retaining efficiency.

References

[1] Andrews, D.W.K. and J.C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator; *Econometrica*, **60**, 953-66.

[2] Browning, M. and C. Meghir (1991), The effects of male and female labor supply on commodity demands, *Econometrica*, **59**, 925-51.

[3] Burnside, C. and M.S. Eichenbaum (1996), Small-sample properties of GMM-based Wald tests, *Journal of Business and Economic Statistics*, **14**, 294-308.

[4] Cameron, A.C. *et. al.* (1988), A microeconometric model of the demand for health insurance and health care in Australia, *Review of Economic Studies*, **55**, 85-106.

[5] Davidson, R. and J.G. MacKinnon (1993), *Estimation and inference in econometrics*, Oxford University Press.

[6] de Luna, X. and P. Johansson (2001), Testing exogeneity under distributional misspecification, IFAU - Office of Labour Market Policy Evaluation, Uppsala, Sweden, working paper 2001:9.

[7] Deb, P. and A.M. Holmes (2000), Estimates of use and costs of behavioural health care: a comparison of standard and finite mixture models, *Health Economics*, **9**, 475-89.

[8] Gallant (1987), *Nonlinear Statistical Models*, Wiley, New York.

[9] Gurmu, S. (1997), Semi-parametric estimation of hurdle regression models with an application to medicaid utilization, *Journal of Applied Econometrics*, **12**, 225-42.

[10] Hansen, L.P. and J. Heaton (1996), Finite-sample properties of some alternative GMM estimators, *Journal of Business & Economic Statistics*, **14**, 262-281.

[11] Hausman, J.A. (1978), Specification tests in econometrics, *Econometrica*, **46**, 1251-71.

[12] Mullahy, J. (1997), Instrumental-variable estimation of count data models: applications to models of cigarette smoking behavior, *Review of Economics and Statistics*, **79**, 586-93.

[13] Newey, W.K. (1985), Maximum likelihood specification testing and conditional moment tests, *Econometrica*, **53**, 1047-70.

[14] Newey, W.K. and K.D. West (1987a), Hypothesis testing with efficient method of moments estimation, *International Economic Review*, **28**, 777-787.

[15] Newey, W.K. and K.D. West (1987b), A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, **55**, 703-08.

[16] Pohlmeier, W. and V. Ulrich (1995), An econometric model of the two-part decision-making process in the demand for medical care, *Journal of Human Resources*, **30**, 339-61.

[17] Romeu, A. and A.M. Vera-Hernández, A semi-nonparametric estimator for counts with an endogenous dummy variable, mimeo, http://pareto.uab.es/wp/2000/45200.pdf

[18] Ruud, P.A. (1984), Tests of specification in econometrics, *Econometric Reviews*, **3,** 211-42.

[19] Terza, J.V. (1998), Estimating count data models with endogenous switching: sample selection and endogenous treatment effects, *Journal of Econometrics*, **84**, 129-54.

[20] Van Ophem, H. (2000), Modeling selectivity in count-data models, *Journal of Business and Economic Statistics*, **18**, 503-511.

[21] Vera-Hernández, A.M. (1999), Duplicate coverage and demand for health care. The case of Catalonia, *Health Economics*, **8**, 579-98.

[22] Windmeijer, F.A.G. and J.M.C. Santos Silva (1997), Endogeneity in count data models: an application to demand for health care, *Journal of Applied Econometrics*, **12**, 281-94.

Tables

Table 1: Frequency of Rejections, Linear Model, True Null Hypothesis

	10%	5%	1%
H0(f)	0.0622	0.0431	0.0229
H0(s)	0.2467	0.1798	0.0993
CRIT	0.1064	0.0489	0.0081
H1(f)	0.0435	0.0157	0.0014
H1(fbe)	0.0970	0.0435	0.0071
H1(s)	0.0843	0.0354	0.0032
H1(sbe)	0.1049	0.0494	0.0087
H2(f)	0.0282	0.0117	0.0012
H2(fbe)	0.0970	0.0435	0.0071
H2(s)	0.0990	0.0435	0.0065
H2(sbe)	0.1122	0.0580	0.0108

Table 2: Frequency of Rejections, Linear Model, False Null Hypothesis

	10%	5%	1%
H0(f)	0.1166	0.0834	0.0442
H0(s)	0.4356	0.3631	0.2439
CRIT	0.3123	0.1957	0.0552
H1(f)	0.1995	0.1017	0.0190
H1(fbe)	0.2992	0.1841	0.0481
H1(s)	0.3577	0.2170	0.0483
H1(sbe)	0.3881	0.2606	0.0851
H2(f)	0.1852	0.0974	0.0213
H2(fbe)	0.2992	0.1841	0.0481
H2(s)	0.3607	0.2290	0.0567
H2(sbe)	0.3890	0.2635	0.0907

Table 3: Frequency of Rejections, Count Model, True Null Hypothesis

	10%	5%	1%
H0(f)	0.0946	0.0766	0.0567
H0(s)	0.2094	0.1479	0.0743
CRIT	0.1911	0.1168	0.0412
H1(f)	0.2096	0.1438	0.0685
H1(fbe)	0.1319	0.0781	0.0326
H1(s)	0.1168	0.0617	0.0166
H1(sbe)	0.1164	0.0634	0.0208
H2(f)	0.1911	0.1277	0.0576
H2(fbe)	0.1095	0.0611	0.0236
H2(s)	0.1108	0.0565	0.0158
H2(sbe)	0.0913	0.0422	0.0103

Table 4: Frequency of Rejections, Count Model, False Null Hypothesis

	10%	5%	1%
H0(f)	0.3495	0.3141	0.2560
H0(s)	0.4692	0.4087	0.3021
CRIT	0.6820	0.5664	0.3438
H1(f)	0.8236	0.7698	0.6499
H1(fbe)	0.4280	0.2761	0.0902
H1(s)	0.5663	0.4610	0.2588
H1(sbe)	0.3870	0.2409	0.0724
H2(f)	0.8312	0.7739	0.6520
H2(fbe)	0.4038	0.2464	0.0743
H2(s)	0.5653	0.4558	0.2569
H2(sbe)	0.3897	0.2312	0.0525

Table 5: *p*-value of GMM Specification Test, NLIV Estimator

	Men	Women
OBDV	0.758	0.738
OPV	0.419	0.161
IPV	0.119	0.802
ERV	0.934	0.073
DV	0.347	0.374
PRESCR	0.731	0.588

Table 6: Specification test p-values, Men

	OBDV	OPV	IPV	ERV	DV	PRESCR
H0(f)	1.000	1.000	1.000	0.988	0.000	1.000
H0(s)	0.000	0.666	0.005	0.411	0.003	0.000
CRIT	0.000	0.935	0.008	0.669	0.058	0.000
H1(f)	0.000	0.956	0.002	0.678	0.000	0.000
H1(fbe)	0.000	0.743	0.000	0.000	0.000	0.000
H1(s)	0.000	0.721	0.008	0.487	0.001	0.000
H1(sbe)	0.005	0.736	0.221	0.542	0.174	0.031
H2(f)	0.000	0.980	0.001	0.470	0.000	0.000
H2(fbe)	0.000	0.415	0.000	0.000	0.000	0.000
H2(s)	0.000	0.728	0.008	0.494	0.001	0.000
H2(sbe)	0.006	0.741	0.260	0.550	0.154	0.015

Table 7: Specification test p-values, Women

	OBDV	OPV	IPV	ERV	DV	PRESCR
H0(f)	0.000	1.000	1.000	0.892	0.956	0.000
H0(s)	0.000	1.000	0.468	0.493	0.846	0.000
CRIT	0.000	0.001	0.530	0.036	0.187	0.001
H1(f)	0.000	0.003	0.636	0.383	0.432	0.000
H1(fbe)	0.000	0.000	0.462	0.000	0.070	0.000
H1(s)	0.000	0.149	0.590	0.517	0.855	0.000
H1(sbe)	0.043	0.134	0.684	0.533	0.852	0.048
H2(f)	0.000	0.902	0.501	0.042	0.313	0.000
H2(fbe)	0.000	0.775	0.007	0.000	0.054	0.000
H2(s)	0.000	0.938	0.618	0.527	0.848	0.000
H2(sbe)	0.046	0.945	0.700	0.541	0.844	0.047